DOCUMENT RESUME

ED 340 756 TM 018 012

AUTHOR Ewell, Peter T.; Jones, Dennis P.
TITLE Actions Matter: The Case for Indirect Measures in Assessing Higher Education's Progress on the National Education Goals.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
PUB DATE 7 Oct 91
NOTE 55p.; Commissioned paper prepared for a workshop on Assessing Higher Order Thinking & Communication Skills in College Graduates (Washington, DC, November 17-19, 1991), in support of National Education Goal V, Objective 5. For other workshop papers, see TM 018 009-024.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Academic Achievement; College Faculty; *College Graduates; Communication Skills; *Critical Thinking; Data Collection; *Educational Assessment; Educational Objectives; *Evaluation Methods; Higher Education; National Programs; *National Surveys; Problem Solving; Thinking Skills
IDENTIFIERS America 2000; *Educational Indicators; Indirect Assessment; *National Education Goals 1990; Performance Based Evaluation; Progress Reporting

ABSTRACT

This paper makes a case for the development of several indirect indicators of national progress in reaching the collegiate attainment objective of the National Education Goals of 1990. These indicators would be developed concurrently with a proposed performance-based assessment to be administered to a national sample of graduating college seniors. Indirect approaches will be easier to develop than direct assessments and will make data available at an earlier point. Specific types of indicators include: (1) institutional curricular and skills requirements for attainment of the baccalaureate; (2) indicators of good instructional practice consistent with the development of critical thinking and communication abilities; (3) national transcript studies; (4) faculty teaching practice surveys; and (5) surveys of current and graduating students. The strengths and weaknesses of these approaches are summarized. Because of the special nature of indirect indicators the standard review questions for these papers do not apply. Two figures illustrate the arguments, and an 18-item list of references is included. An appendix lists several existing survey instruments with potential for measurement in indirect assessment. Comments by R. Calfee, E. M. Greenberg, and M. L. Tenopyr on this paper are provided. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED340756

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ACTIONS MATTER: THE CASE FOR INDIRECT MEASURES IN ASSESSING HIGHER EDUCATION'S PROGRESS ON THE NATIONAL EDUCATION GOALS

Peter T. Ewell
Dennis P. Jones

National Center for Higher Education Management Systems, Inc.
P.O. Drawer P
Boulder, CO 80301-9752

October 7, 1991

TM018012

BEST COPY AVAILABLE

ABSTRACT

ACTIONS MATTER: THE CASE FOR INDIRECT MEASURES IN ASSESSING HIGHER EDUCATION'S PROGRESS ON THE NATIONAL EDUCATION GOALS

This paper makes a case for the development of a number of indirect indicators of national progress in reaching the collegiate attainment objective of the National Education Goals. These indicators would be developed concurrently with a proposed performance-based assessment to be administered to a national sample of graduating college seniors. The case rests in essence on two grounds. First, direct assessments of the ability of college graduates to "think critically, communicate effectively, and solve problems" are technically complex and will take many years to develop. Indirect approaches hold the promise of generating useful data at an earlier point, and therefore of allowing higher education to remain at the center of national policy interest. Secondly, the information provided by such indicators--particularly if they are primarily directed at "good practices" in undergraduate instruction whose efficacy in producing desired outcomes has been empirically established--is useful in itself for inducing action and for guiding the further development of national policy.

Specific types of indicators noted include, a) institutional curricular and skills requirements for attainment of the baccalaureate, b) indicators of instructional "good practice" consistent with the development of critical thinking and communications abilities, and c) student behavior and self-reported gains. Types of data sources explored for the collection of indirect indicators include, a) institutional administrative records, b) surveys of institutional practice, c) national transcript studies, d) surveys of faculty teaching practice, and e) surveys of current and graduating students. Each of these methodologies is well-developed, but they have not in most cases been applied to national samples. A summary of strengths and weaknesses of the indirect indicators approach is provided, together with recommendations for next steps should it be decided that this approach is worth further consideration.

Because of the special nature of indirect indicators, the standard review questions proposed for these papers do not completely apply. Regarding the domain of assessment, the best answer is that the domain consists of institutional policies and practices known or presumed to affect the development of higher-order cognitive skills, and participant testimony about both the presence of these factors and their own assessments of the degree to which such outcomes are in fact being attained. Regarding "who is to be tested", the best answer is that valid national samples can be obtained of institutions, graduating student transcripts, faculty members, and current students. Regarding "what standards should be applied?", the clear answer is none--the proposed indicators are intended to be indicative only and are by nature inappropriately conceived as performance standards. Finally, regarding "what instruments and approaches should be used?", the answer is a combination of existing (though not aggregated) institution-level information and surveys of key populations.

ACTIONS MATTER: THE CASE FOR INDIRECT MEASURES IN ASSESSING HIGHER EDUCATION'S PROGRESS ON THE NATIONAL EDUCATION GOALS

Late last summer, a technical resource group convened by the National Education Goals Panel under the leadership of Governor Romer recommended eventual development of "a new kind of assessment . . . of a type similar to the National Assessment of Educational Progress (NAEP) at the College Level" to help track the nation's progress in attaining one of the few objectives under the National Education Goals that explicitly references higher education: the ability of graduating college seniors to "think critically, communicate effectively, and solve problems." In recommending development of this measure, however, the group reminded the Panel that building such a measure would be a long-term and a potentially expensive endeavor. In the short run, therefore, they also recommended that the Panel explore development of additional "indirect" indicators of progress, based largely on the typical educational experiences of college students and on the curricular innovations and changes in instructional practice actually adopted and fielded by the nation's colleges and universities. The purpose of this paper is to make the case for indirect indicators of this kind. In doing so, however, we wish to make our position--consistent with that of the National Goals Panel--completely clear: if it is the nation's intention to adequately assess the higher order cognitive skills of its college graduates, indirect measures will not substitute for the kind of purpose-built, performance-based national assessment now recommended.

Accordingly, the paper is organized as follows. A first section discusses the reasons for advocating the development of a set of indirect indicators, particularly emphasizing the need for short-term, action-oriented measures of progress. A second section outlines the primary characteristics of indirect indicators, and poses some important conceptual choices that must always be made in their development. A third section provides specific examples of indirect indicators of different types, together with a brief analysis of how they might be collected. A fourth section summarizes the particular strengths and weaknesses of indirect indicators as an element of national postsecondary assessment. A final section proposes some next steps that might be undertaken if it is decided that the indirect indicators approach appears promising.

Why Indirect Indicators?

The case for developing indirect indicators to track higher education's progress first rests upon the extreme difficulty of producing direct ones. Past experience in developing meaningful postsecondary general assessments of the kind recommended by the National Goals Panel (most notably the New Jersey "General Intellectual Skills" (GIS) examination) suggests that a timeline of at least five years will be required (Educational Testing Service 1989). As noted by the Technical Panel, a "consensus-building" process will first be needed to translate stated abilities such as "critical thinking" or "problem-solving" into operational terms; currently, such abilities are conceptually ill-specified, and have been operationally reflected in quite different kinds of assessments. At the same time, though experience in the development of large-scale performance-based assessments exists outside the K-12 arena in the form of the GIS and the Adult Literacy scales of the National Assessment of

Educational Progress (NAEP), the technical properties of such assessments are complex and in many cases unknown; standard validity and reliability measures are hard to apply and the results are often subject to unknown and uncontrollable biases. Growing experience in assessing college student populations, moreover, suggests that motivating students to participate and to do their best may be a major problem. All these challenges, no doubt, can be overcome, but the process will take time. Meanwhile, if information about higher education is to inform national policy, it must be collected by other means.

At the same time, there are compelling reasons why indirect indicators may be useful in their own right. Most importantly, such statistics, if constructed properly, can be tied directly to policy action. Indeed, one historical drawback to cognitive assessment results--particularly in postsecondary education--has been the difficulty of linking them to actual aspects of prior educational experience that can be effectively manipulated through policy action (Ewell 1991). Valid test scores can indicate rather precisely what has been accomplished and where deficiencies exist, but they often provide little guidance about what can or should be done. And this appears particularly to be the case for the kinds of higher order abilities noted in the national goals. Substantial research and experience has shown that as the complexity and generality of assessed skills increases, so does the difficulty of showing real growth or of establishing causality (Pace 1979; Pascarella and Terenzini 1991). As a result, it is unlikely that linkages will quickly be established between results on any national outcomes measure in postsecondary education and what can and ought to be done to improve practice.

This context suggests two major reasons to pursue the development of indirect indicators. First, they provide important additional information with which to make sense of the findings of end-result assessments. Except as a pure benchmark of progress, it makes little policy sense to collect outcomes information in the absence of information on key processes that are presumed to contribute to the result. Higher education, moreover, is particularly in need of information about contexts and processes because of its immense variety. Not only do colleges and universities exist in many forms with diverse educational missions, but unlike the situation typical of K-12 education, there is little curricular commonality in the experiences of college students; even at the same institution, because of wide curricular choices, no two students are likely to have experienced the same behavioral "program." As noted by many assessment observers, therefore, information on outcomes alone is virtually uninterpretable in the absence of information about key experiences (Astin 1991). Secondly, indicators tied to key instructional practices provide clear policy leverage for action. A major difficulty of outcome-based performance-funding experiments in higher education has been relative lack of direction about where to invest to obtain best results. Funding approaches targeted directly to particular activities, such as general education reform or minority achievement in the form of challenge or incentive grants, have generally proven more effective than performance incentives in college and university settings. This is largely because of a greater need to focus greater initial attention on consistent action and innovative processes in higher education than in the K-12 context. Indeed, past experience with colleges and universities suggests that national indicators linked to the particular attributes of an undergraduate education that prior research has already demonstrated have an impact on achievement

might prove more effective in leveraging improvement than a set of indicators anchored exclusively in postsecondary outcomes.

Concern with processes as well as outcomes indicators is also a prominent feature of the "quality movement" in business and industry, and increasingly, approaches to reform based on "total quality management" (TQM) are being discussed in higher education as well. Such approaches, it is important to stress, do not dispense with outcomes information as is sometimes claimed. Indeed, periodically monitoring information on product sales, on numbers of defects, and on customer satisfaction is critical to the TQM approach as practiced in industrial settings. Equally important, however, are "continuous monitoring" indicators that allow managers to much more clearly understand and manipulate the processes that they manage. Analogously, both kinds of indicators are important in improving higher education. Outcomes indicators of the type recommended by the National Goals Panel are arguably needed to provide a consistent "market signal" about the level of quality produced by the nation's higher education system. But additional national data are needed on key higher education processes to actually guide the process of improvement and to help direct available resources to where they are most needed.

What Exactly Is an "Indirect Indicator?"

An initial difficulty associated with proposing the development of "indirect indicators" of collegiate attainment is conceptual. While cognitive indicators at least at first appear straightforward, many different kinds of statistics have been proposed as "indirect" measures of academic progress. Before describing explicitly the kinds of indicators that might be developed,

therefore, it is useful to briefly note some distinctions between the quite different types of statistics that might be pursued. At the same time, it must be emphasized that all indicators of educational attainment are in some sense indirect. Purpose-built, performance-based assessments of particular areas of knowledge and skill, such as that proposed by the National Goals Panel for tracking progress in collegiate attainment, are no exception and like the results of any test, should not be confused with the actual entity that they purport to represent. Differences between the properties of "direct" and "indirect" measures noted in the balance of this paper, therefore, should be read more as differences in degree than as differences in kind. Nevertheless, two conceptual distinctions appear important in discussing them.

A first distinction rests on the degree to which the proposed indicator is causally related to the core phenomenon being examined. Useful and reliable indicators for "tracking progress" regarding educational attainment, for instance, might be developed which are highly correlated with cognitive abilities, but which are not themselves directly related to the development of these abilities. "Proxy" indicators of this kind are commonly used in other fields to document overall trends, but their major drawback is that they cannot generally be used to guide policy. Attempting to use them in this manner, indeed, may be actively harmful, as scarce managerial attention is expended toward maximizing indicator values in themselves, rather than on making more fundamental improvements. This has been a major externality of developing high-stakes indicator systems--direct or indirect--at all levels of education.

A second distinction rests on the degree to which the proposed indicator focuses on the outcome itself or instead on the processes and investments that are critical for its attainment. Past (and certainly founded) criticisms of assertions about the quality of higher education note that the basis for such judgments rested largely on inputs and processes, while actual attainments remained unexamined. The result has been much greater attention to the outcomes of higher education over the past five years, both at the state level and in accreditation. At the same time, the use of input and process information in these arenas has not disappeared--largely because they are useful in guiding actual improvement.

Figure 1 embodies these two primary distinctions in a classification of the kinds of indicators that might be considered in assessing national collegiate attainment. Its horizontal dimension embodies the first distinction between direct and "proxy" measures, while its vertical dimension incorporates the second distinction between outcomes and process domains. Note, however, that "process" indicators occur in <u>two places</u> in this representation--as "proxy" outcomes measures and as direct measures of important phenomena of policy interest in themselves.

Following the logic of this scheme, the proposed assessment of graduating college seniors' ability to "think critically, communicate effectively, and solve problems" is shown in the upper left-hand corner. As noted, considerable effort will be required to develop such a measure, both because of its complex (and unknown) technical properties and because of the amount of consensus-building required to adequately define the abilities being assessed. But a number of "proxy" indicators of these abilities can be considered. For

"critical thinking" or "effective communications" abilities among college graduates, for instance, trends in existing assessment results, though highly flawed, can serve this role. Most prominent among these applications have been results on the Graduate Record Examinations (GRE) and the Adult Literacy component of the National Assessment of Educational Progress (NAEP). Past meta-studies have used such results indicatively to suggest trends in more complex abilities, and have been particularly careful to limit their conclusions to the bounds of the instruments used (Adelman 1985). A second example--even more indirect--is to track patterns in the typical standards used to certify attainment to indicate trends in actual levels of attainment. A recent example is provided by a National Endowment for the Humanities report, whose analytical basis is an analysis of the content of current national examinations, rather than their results (Cheyney 1991). Finally, self-reports of college graduates regarding their own abilities and current behaviors can be collected as indicative of actual underlying student abilities.

Potential instructional process indicators can also be distinguished on the basis of the directness with which the phenomena of interest are reflected. Among the most direct are statistics on typical behaviors. Several recent institutional and multi-institutional studies, for instance, have investigated the prevalence of innovative classroom practices in typical undergraduate instructional settings--for instance, the frequency with which students are presented with material that requires the kinds of active engagement needed for developing critical thinking skills (Chickering and Gamson 1987; Light 1990). Others have examined patterns of typical undergraduate course-taking to determine the degree to which graduating college students are in fact

exposed to particular bodies of material (Ratcliffe and Associates 1990. Zemsky 1989). Indirect counterparts for these sources of evidence include patterns of institutional resource utilization consistent with "good practice" and student self-reports about their own study habits and learning practices.

Although statistics about instructional processes on a national basis (direct or indirect) might be useful for many purposes, what in particular is required for them to serve as reliable indicators of actual collegiate attainment? First, like any "proxy" indicator, trends in the incidence of such processes must be demonstrably related to changes in resulting student attainments. This will require both active piloting of any proposed statistics using multi-institutional samples, and extensive background research in the empirical literature on college student learning in order to determine the direction and strength of any previously established associations. Second, it must be shown that such statistics can be meaningfully collected across a wide range of institutional types without becoming so "noisy" that their underlying value is lost. National transcript studies of graduating college seniors, for instance, have generally focused on content classifications for determining patterns of coursetaking, based on established disciplines; determining exposure to "critical thinking" may eventually require as much preliminary consensus-building about course coverage as would the development of a more direct measure of college student attainment. Finally, as noted earlier, the resulting indicator should not inadvertently induce institutions or policymakers to behave inappropriately merely to maximize the value of the indicator. Each proposed indicator of this kind should thus be subject to critical review and careful field testing to determine the degree to which it might deliberately be manipulated.

What Specific Kinds of Indicators Might be Investigated?

With the caveats noted, a number of promising indirect indicators of collegiate attainment might usefully be pursued in support of the National Educational Goals. Few are currently collected on a national basis, but most could be addressed through use of a known instrument, methodology, or record-keeping procedure. In many cases, in fact, multiple data-collection procedures could be used; for instance, the average amount of student writing required in a typical undergraduate course of study might be obtained both through an analysis of a national sample of transcripts together with a syllabus review, or by means of self-reports on a national survey of graduating students. Accordingly, this section of the paper consists of two distinct parts. First, major substantive classes of indirect indicators are reviewed. Secondly, the principal methods available for collecting such data are noted. In addition, both dimensions are summarized in figure 2, and several examples drawn from existing instruments are noted in Appendix A.

Primary Types of Indicators

Useful indirect indicators for undergraduate attainment are of quite different types. In general, however, they embrace what colleges and universities require of their students, what typically happens in a collegiate classroom or course of study, and what college students do as a part of their education. Using this logic, the following types of indicators might be usefully developed:

Institutional Requirements. Indicators of this type address the degree to which institutional curricular requirements contain features expected to be associated with collegiate attainment in the areas of critical thinking, communications or problem-solving. Examples include:

- specific proficiencies required for attainment of the baccalaureate degree--for example, explicit demonstrations of writing, speaking, computational ability, foreign language proficiency, etc.

- specific types of experiences required for attainment of the baccalaureate degree--for example, is it possible for students to graduate without having written a major research paper, taken a math course, taken a laboratory science or taken a foreign language?

- specific "capstone" or other integrative experiences required for graduation--for example, an internship, problem-oriented senior seminar, or senior thesis or project.

Instructional "Good Practice." Indicators of this type address the degree to which typical student instructional experiences are consistent with established principles of good practice in undergraduate teaching (Chickering and Gamson 1987)--for instance "active learning", frequent "feedback" on performance, or frequent student/faculty contact. Examples include:

- typical class-sizes encountered in lower-division courses--for example, how likely is it that a lower-division student (or first-term freshman) is

enrolled in at least one class containing fifteen or fewer students, in which "active participation" is likely?

- instructional experiences reported by students as typical of their undergraduate coursework--for example, the frequency of writing or speaking required, levels of participation in group study or explicit problem-solving experiences, the amount and type of out-of-class work required, numbers of assignments requiring outside independent work, or the proportion of course final examinations taken that required an essay or problem-solving component.

- additional out-of-class experiences reported by students as typical of their undergraduate experience--for example, frequency of out-class-contact with faculty members, active participation in faculty research projects, participation in on- or off-campus work related to their course of study, participation in group study, frequency of independent college-related research or study, or frequency of tutoring another student.

- institutional policies or investments consistent with "good practice" in undergraduate instruction--for example, the proportion of lower-division sections (or student credit-hours) taught by full-time or senior faculty members, the proportion of classes taught by graduate teaching assistants, average size of classes in which freshman students are enrolled, existence of policies or criteria for faculty promotion and tenure emphasizing excellence in undergraduate teaching, or existence and coverage of institutional programs and policies regarding the assessment of student learning.

Student Behaviors and Self-Reported Gains. Indicators of this type address the degree to which students themselves report behaviors and outcomes consistent with good practice in undergraduate instruction--and particularly in the acquisition of critical thinking, communications, or problem-solving skills. Examples include:

- student use of time ("time on task") in selected areas--for example, reading, writing, working mathematical or scientific problems, talking in class, talking with other students about class-related material, or working on independent research or library assignments.

- student self-reported gains in selected areas--for example analytical/problem-solving skills, oral and written communications skills, ability to think critically, or ability to work cooperatively.

- student self-reports regarding their reactions to college-level work--for example, the proportion of current college students reporting being actively challenged by their classes or by out-of-class assignments, or the level of self-reported interest and involvement in instruction when compared to job-related activities or extracurricular work.

Each of these possible types of indicators requires further study regarding its feasibility, and the inventory above is far from exhaustive. And it is again important to emphasize that like indirect measures in any area, statistics such as these are best used in combination. If pursued as part of a national reporting strategy, state and institutional experience suggests, at least ten to twelve such indicators of different types should be developed

and, as noted, they should be used both to supplement the results of a more direct measure of college student attainment and to provide an important near-term indication of higher education's progress.

Primary Sources of Indicator Data

Data sources for the types of indirect indicators noted above are of many kinds. Most of the appropriate instruments and methods have been extensively used at the institutional level and occasionally at the state level, but few are currently in place at the national level. Modifying them for use with national samples, however, would in most cases not be difficult and would certainly entail far less time and cost than would the development of a direct assessment of collegiate learning. Currently, the following kinds of data sources appear available and appropriate.

Institutional Administrative Records. Colleges and universities currently keep extensive records on instructional activity and resource utilization. If standard indicator definitions could be developed, these data might be tapped annually to help broadly determine how institutions are deploying their available resources in support of effective undergraduate education. Examples of the kinds of indicators that might be produced from base statistics of this kind include, the average size of key classes, the proportion of lower-division classes taught by graduate students and full-time faculty, and the proportion of small classes typically taken by first-year students. Such data might be collected on the basis of a national cross-sectional sample of institutions. Alternatively, key statistics might eventually be made a part

of institutional reporting to the National Center for Education Statistics (NCES) under the Integrated Postsecondary Education Data System (IPEDS).

Surveys of Institutional Practice. Institutional surveys have in the past been used by higher education scholars to determine the degree to which colleges and universities are engaged in innovative practices in undergraduate instruction. Prominent examples here include a study on undergraduate reform recently undertaken by the National Center for Research on Postsecondary Teaching and Learning (NCRPTAL) at the University of Michigan (Peterson 1987) and an ongoing "Registry of Undergraduate Reform" maintained by the California State University (Vandament 1990). Similar are the occasional "quick response" surveys on various topics conducted by NCES. Such instruments typically rely on defined institutional respondents, selected by position, to report institutional activities, and generally a sample of institutions is surveyed. Examples of the kinds of indicators that might be obtained by means of this method include minimum skills and curriculum-coverage requirements for receipt of the baccalaureate, particular curricular emphases on critical thinking, communication and problem-solving, and institutional activities in the assessment of student learning.

National Transcript Studies. Transcript records contain information on typical college coursetaking patterns, and can be used to determine the degree to which students are generally exposed to particular bodies of material. Several national studies of this kind have been undertaken for different purposes (for example, Zemsky 1989, Ratcliffe and Associates 1990). Such studies usually examine coursetaking patterns on a discipline basis--addressing questions such as the number and proportion of courses in a total

baccalaureate career taken in key identified areas (for instance, science, math, or history), or the overall "concentration" and "cohesiveness" of the curriculum (Zemsky 1989). Though challenging, similar methodologies based on a national sample might be developed that, together with supplied course descriptions and required assignments, could suggest the frequency with which students are typically graduating after having completed certain key assignments or experiences--for example an independent research project or a senior "capstone" experience.

Surveys of Faculty Teaching Practice. Higher education researchers have also developed many surveys of college and university faculty. While most are directed at issues of career, compensation, and scholarly research, several also contain items on teaching practice. Useful examples of such surveys currently include the faculty "self-assessment" instrument developed as part of the Wingspread "Seven Principles of Good Practice in Undergraduate Instruction" (Gamson and Poulsen 1989), and the "Faculty Survey" periodically administered to a national sample by the Higher Education Research Institute of the University of California at Los Angeles. Examples of the kinds of indicators that might be developed using this method are the proportion of faculty time spent working with students, the reported incidence of "active learning" activities used by faculty members in the classroom, levels of student participation in independent or faculty research work reported by faculty, and levels of faculty participation in professional development activities directed toward the improvement of teaching. Surveys of this kind can also use faculty as "expert witnesses" regarding typical institutional practices, outcomes, and reward structures. The UCLA faculty survey, for instance, asks respondents to rate such items as the frequency of student-

faculty contact, the importance in instruction of various intended outcomes of college, and the typical types of evaluation methods used to assess student classroom performance (see Appendix A).

Surveys of Current and Graduating Students. Questionnaires administered to current and former college students are typically used by individual colleges and universities to determine levels of satisfaction with instruction and other services, patterns of typical student activities, and self-reported outcomes of instruction. Several of these are administered on a national basis. Among the most prominent are the freshman and follow-up surveys of the Cooperative Institutional Research Program (CIRP) administered annually to representative national samples, and the College Student Experiences Questionnaire (CSEQ) widely used in assessment activities in many types of institutions; both are housed at the Higher Education Research Institute of the University of California at Los Angeles. Examples of the kinds of items contained in these questionnaires include reported levels of participation in activities such as internships or faculty research projects, time spent in various activities (for example, group study or tutoring another student), frequency of student-faculty and group interaction, self-reported class content, and self-reported gains on a range of outcomes. Several items contained on these questionnaires are potentially usable as indicators without change; for the future, enhancements might be made in both item content and the sampling base for such instruments for use in a national indicators system (see Appendix A).

It is important to recognize that this diversity of data sources is itself important in developing a reliable set of indicators. Indeed, the nature of

indirect indicators is such that confidence in the data increases as the number of visibly distinct data sources tapped also grows. As summarized in figure 2, moreover, many of the same types of indicators can and should be collected by means of several different sources. As a result, it is recommended that all or most of these sources be explored, and that at least three be used concurrently in any proposed national system.

What are the Strengths and Weaknesses of this Approach?

In discussing the viability of developing indirect indicators of collegiate attainment, the explicit strengths and weaknesses of this approach in comparison with direct assessment must be considered. At least five major trade-offs are apparent and should be carefully examined. Briefly, they are as follows:

1. Indirect indicators are straightforward and cheap, but more than one will be required. Current estimates of the time required to develop an authentic, performance-based national assessment of collegiate attainment are in the neighborhood of five to eight years, and potential development costs are very high. Indirect indicators, in contrast, are relatively easy to conceptualize and relatively cheap to collect. But to be credible on a national basis, experience suggests, multiple indirect indicators will be needed, relying on quite different sources of data. As a result, considerable initial developmental work will be needed to determine the appropriateness of an indirect indicators approach, and the feasibility of an integrated national data-collection design to support this approach. Costs and timelines for such an activity, however, still appear on balance

to be far below those required for a direct assessment of collegiate attainment and this approach would allow some useful national statistics to be produced at an earlier point than now anticipated.

2. Indirect indicators are relatively easy to collect, but are correspondingly hard to interpret. While indirect indicators are compellingly concrete, they at best provide hints and guesses of the degree to which college graduates actually possess the abilities of interest. As noted, considerable background work in the research literature on college impact will be required to establish minimum levels of correlation or causal connection between increases in such indicators and actual gains in critical thinking, communications, or problem-solving abilities. Conclusions drawn from such indicators will be highly inferential--even when they are presented in combination--and should be treated with caution. Although they can and should be developed more quickly than direct assessments, such statistics should not be seen as a substitute for the kinds of measures proposed for eventual development by the National Goals Panel. In the long run, they should enhance these measures, and in the short run they can provide a critical early source of information about higher education's process.

3. Indirect indicators provide excellent policy leverage for changing instructional practice, but they may also lack overall public credibility as indicators of national progress. A major strength of many of the kinds of statistics discussed above is their explicit focus on entities and actions over which faculty, institutional leaders, and state policymakers have a measure of control. Past experience has demonstrated that it is

generally far easier to provide incentives centered on concrete policies and practices than to reward "outcomes"--particularly when these are at a high level of generality and are difficult to relate to particular patterns of instruction or student experience. Many commentators in higher education have noted that the primary problems associated with greater achievement do not include a lack of knowledge about what works. Instead, the greater barriers are lack of systemic incentives to put in place the kinds of practices that established research already suggests are effective (Stark and Lowther 1986; McKeachie et al. 1986). A national indicator system directed toward "good practice" might help redress this balance, and would be consistent with what appears to be the "next step" in the National Education Goals process. As incoming chair of the National Governors' Association, for instance, Governor Ashcroft is proposing a "Lifelong Learning Action Team" that among other things, will explore "how to shift the campus focus to undergraduate education, student outcomes, and the quality of instruction and curriculum" (national Governors' Association 1991). Indirect indicators of the kinds proposed are highly consistent with this initiative and would explicitly respond to emerging complaints about deficient performance in colleges and universities--most of which are centered on processes. But because they are oriented toward inputs and processes rather than outcomes, such indicators may also lack general public credibility for reporting progress in goal attainment. For better or worse, results obtained through more direct assessments (particularly standardized examinations) have become widely accepted as the true public criterion of success in education, and this will be a condition not easily reversed.

4. Indirect indicators may help individual colleges and universities improve what they do, but they are less able to provide a clear "focus of energy" for mobilizing public support for improvement on a national basis. Statistics of the kinds noted above have potential value in helping individual institutions to allocate available resources in support of better instructional practices. But because they are hard to more generally interpret, and because a relatively large number of independent indicators will be required to make valid inferences about progress, such indicators are less likely than a single, widely-publicized, outcomes statistic to serve effectively as a rallying point for inducing action. Certainly, a major function of the National Education Goals process itself is to do exactly that, and a primary reason for embarking on assessment in connection with the process in the first place is to establish simple benchmarks that can serve as a focus for subsequent energy and action. Past experience with indirect indicators suggests that they cannot be easily or usefully combined into a single summary statistic that can serve this purpose without substantial distortion and loss of validity. As a result, their use may be inherently limited in sustaining public attention and in mobilizing political action.

5. Indirect indicators will in the short run likely prove more palatable to the academy than will more direct assessments of collegiate attainment, but in the long run their collection and use may encounter broader resistance from colleges and universities. Initial institutional and faculty resistance has been a hallmark of the national higher education assessment movement since it began some seven years ago. For the most part, resistance has been based on three grounds--lack of validity on the

part of the instruments used, fear of inappropriate comparisons among programs and institutions with differing goals and clienteles, and the cost and effort required to collect such information. It is likely that similar objections will also be raised against any national attempt to directly assess collegiate outcomes, and that indirect indicators initially will be preferred. Indeed, institutional experience with assessment strongly suggests that faculty will find such evidence as survey results and analyses of coursetaking patterns far more useful than test scores (Banta and Moffett 1987). But such indicators also have the potential to much more directly reveal, and therefore influence, what faculty and institutions actually do. And as a result, they are likely to encounter increasing resistance as external assessment devices, because their consequences for changing existing practice are far more apparent. Institutional experience with assessment strongly suggests that this will eventually occur. Resistance to the Minnesota State University System's "Q-7" initiative, for instance, has to date been more centered on its apparent elements of "curricular prescription" than on its outcomes components, just as Florida's "Gordon Rule" requiring specified amounts of writing be incorporated in freshman courses has proven at least as controversial as the state's "rising junior" examination system. Similarly, many faculty appear ready to fall back on standardized outcomes examinations as the basis for institutional assessment, when confronted with alternatives that would require them to change what they do in the classroom. These experiences suggest the potential of growing resistance to national indicators focused on "good practice" in higher education as the probable implications of this process become clear.

What Should Happen Next?

The types of indirect indicators noted in this paper are intended only as examples of what might eventually be developed. They are not advanced as concrete proposals for collecting or reporting any given statistic. Considerable background work will be required in order to determine the feasibility and relative value of this approach in tracking progress on attainment of the National Education Goals--or indeed, for any other national purpose. If the approach in general is considered promising, appropriate background developmental work should proceed immediately. Requirements for further development, if a decision is made to proceed, might include the following activities:

- a systematic review of available national data collection instruments and methodologies that might within one year be capable of generating one or more indicators. As noted, several extant data systems of various types are currently in place at the national level or are based on valid national samples; they include the CIRP, several national transcript studies, and institutional statistics collected by NCES. These sources should be examined to determine, a) if the information they provide is adequate for any proposed reporting purpose, b) if the sampling basis allows valid generalizations to national populations, and c) if straightforward modifications to these procedures might quickly be made to allow them to be used for national reporting purposes.

- a major background paper on the validity of indirect indicators in assessing progress in collegiate attainment. This paper would consist primarily of a

systematic review of the research literature on college outcomes to identify causal and correlational linkages between possible "proxy" and process indicators and actual gains in students' ability to "think critically, communicate effectively, and solve problems." The paper would result in a set of conclusions about cost-effective approaches to obtaining valid and appropriate indicator data linked directly to desired postsecondary outcomes.

- a <u>feasibility study</u> to determine the costs and logistics associated with collecting information of this kind from a typical sample of institutions and potential respondents. While in some cases, instruments and methodologies for collecting items of interest have been developed and used on national samples, in most cases they have not. Accordingly, the major goals of a feasibility study should be first to develop an appropriate set of draft data collection instruments or protocols, and then to pilot these approaches on a reasonable sample of institutions or respondents. Alternatives for such a study might involve, a) proceeding in conjunction with an ongoing national study (for instance, the current longitudinal study of the National Center on Postsecondary Teaching, Learning and Assessment at the Pennsylvania State University, or the "General Education" study of the Higher Education Research Institute at the University of California at Los Angeles), or b) selecting all institutions in a single state that already possesses statewide data on higher education outcomes (for example, Florida, Tennessee, or New Jersey). Objectives of the pilot would be to assess the feasibility of collecting the indirect measures themselves, and of establishing empirical links between obtained indicators data and any available data on postsecondary outcomes.

At minimum, this will require a two-year development effort before appropriate national indicators are in place. But again, this timeline is far shorter than that required to develop a direct assessment of collegiate attainment.

In essence, the case for developing indirect indicators of collegiate attainment to track progress on the National Education Goals rests on two grounds. First, direct assessments of the ability of college graduates to "think critically, communicate effectively and solve problems" are technically daunting and will be a long time in coming. Indirect approaches hold the promise of generating useful data at an earlier point, and therefore of allowing postsecondary education to remain within the domain of national public policy attention. Secondly, the information provided by such indicators--particularly if they are primarily directed at "good practices" in undergraduate instruction whose efficacy in producing desired outcomes has been empirically established--is of utility in itself in guiding the development of national policy; such indicators are generally of greater value in inducing institutional change than are outcomes indicators used alone. On balance, it seems a path worth exploring further.

Figure 1

Some Possible Indicators of College Student Attainment:
A Basic Typology

Indicator Domain \ Nature of Indicator	Direct	Indirect ("Proxy)
Outcomes	"National Assessment of Collegiate Performance" (Proposed)	- Existing Assessment Results - Prevailing "Standards" - Graduate Self-Reports *[- "Key Process" Indicators]
*Processes	- Typical Instructional Practices - Typical Curricular Behavior (e.g., Coursetaking Patterns)	- Institutional Resource Utilization - Student Self-Reports

* Note: "Process" Indicators are also "Indirect Outcomes" Indicators in this conception.

Figure 2

Overall Coverage of Indicator Types and Sources of Data

	Sources of Data				
	Institutional Administrative Records	Surveys of Institutional Practice	National Transcript Studies	Surveys of Faculty Teaching Practice	Surveys of Current and Graduating Students
Institutional Requirements		- Curricular Requirements - Requirements for Graduation	- Coursetaking Patterns by Discipline - Exposure to Key Experiences (e.g., capstone course, research project)	- Amount of Writing, Oral Presentation, or Independent Work Required - Types of Exams & Exercises Used to Evaluate Students	- Reported Requirements for Writing, Indep. Projects, etc.
Instructional "Good Practice"	- Levels of Investment (e.g., "Front Loading")	- Innovative Teaching Practices - Assessment Activities		- Use of Instructional Practices (e.g., Group Projects, Independent Projects) - Reported Levels of Institutional Support for Good Practice	- Reported Incidence of Selected Classroom Practices - Assessment of Campus Learning Environment
Student Behavior & Self-Reported Gains		[- Reported Outcomes in Critical Thinking, Communications & Problem-Solving]		- Reported Student Activities (e.g., Time on Task, Independent Work) - Reported Outcomes in Critical Thinking, Communications, and Problem-Solving	- Reported Behaviors (e.g., "Time on Task", Group Study, etc.) - Reported Outcomes in Critical Thinking, Communications, & Problem-Solving

REFERENCES

Adelman, C. (1985). The Standardized Test Scores of College Graduates 1964-1982. Washington, D.C.: U.S. Government Printing Office.

Astin, A.W. (1991). Assessment for Excellence: The Philosophy and Practice of Assessment and Evaluation in Higher Education. San Francisco: Jossey-Bass.

Banta, T.W. and Moffett, M.S. (1987). Performance Funding in Tennessee: Stimulus for Program Improvement. In D.F. Halpern (ed), Student Outcomes Assessment: What Institutions Stand to Gain (New Directions in Higher Education No. 59, pp.35-44). San Francisco: Jossey-Bass.

Cheyney, L.V. (1991). National Tests: What Other Countries Expect Their Students to Know. Washington, D.C.: National Endowment for the Humanities.

Chickering, A.W. and Gamson, Z.F. (1987). Seven Principles for Good Practice in Undergraduate Education. AAHE Bulletin, 39(7), 3-7.

Educational Testing Service. (1989). New Jersey College Outcomes Evaluation Program: A Report on the Development of the General Intellectual Skills Assessment. Princeton, N.J.: Author.

Ewell, P.T. (1991). To Capture the Ineffable: New Forms of Assessment in Higher Education. Review of Research in Education, 17, Washington D.C.: American Educational Research Association, 75-126.

Gamson, Z.F. and Poulsen, S.J. (1989). Inventories of Good Practice: The Next Step for the Seven Principles for Good Practice in Undergraduate Education. AAHE Bulletin, 42, 7-8.

Light, R.J. (1990). The Harvard Assessment Seminars: Explorations with Students and Faculty About Teaching, Learning, and Student Life. Cambridge, MA: Harvard Graduate School of Education and Kennedy School of Government.

McKeachie, W.J., Pintrich, P.R., Lin, Y.G., and Smith, D.A.F. (1986). Teaching and Learning in the College Classroom: A Review of the Research Literature. Ann Arbor, MI: National Center for Research on Postsecondary Teaching and Learning (NCRPTAL), University of Michigan.

National Governors Association. (1991). Redefining the Possible: Achieving the National Education Goals. The Chairman's Agenda for 1992. Washington, D.C.: Author.

Pace, C.R. (1979). Measuring the Outcomes of College: Fifty Years of Findings and Recommendations for the Future. San Francisco: Jossey-Bass.

Pascarella, E.T., and Terenzini, P.T. (1991). How College Affects Students: Findings and Insights from Twenty Years of Research. San Francisco: Jossey-Bass.

Peterson, M.W. (1987). Academic Management Proactices Survey for the Research Program on the Organizational Context for Teaching and Learning. Ann Arbor,

MI: National Center for Research on Postsecondary Teaching and Learning (NCRPTAL), University of Michigan.

Ratcliffe, J.L. and Associates. (1990). Determining the Effect of Different Coursework Patterns on the General Learned Abilities of College Students. Research Institute for Studies in Education at Iowa State University and Center for the Study of Higher Education at the Pennsylvania State University.

Stark, J.S. and Lowther, M.A. (1986). Designing the Learning Plan: A Review of Research and Theory Related to College Curricula. Ann Arbor, MI: National Center for Research on Postsecondary Teaching and Learning (NCRPTAL), University of Michigan.

Vandament, W.E. (1990). Registry of Higher Education Reform. Los Angeles: Council for the Development of Education.

Zemsky, R. (1989). Structure and Coherence: Measuring the Undergraduate Curriculum. Philadelphia: Institute for Research on Higher Education, the University of Pennsylvania.

APPENDIX A

As noted in the body of this paper, several existing survey instruments contain items of potential utility as indirect indicators of "good practice" in undergraduate instruction or of collegiate attainment. This appendix contains a brief review of these items for reference. Specific item texts are not provided, but general item content is noted for each instrument. In each case, the instruments are administered to national samples, but only the CIRP is based on a sampling frame specifically tailored for national reporting; the CIRP, moreover, has resulted in twenty years of historical data that might serve as a useful baseline for charting future progress.

Cooperative Institutional Research Program (CIRP)

The CIRP follow-up is regularly administered to national samples of college students, with many of its items linked to an annual freshman survey so that estimates of change can be made. Among the items of interest on the follow-up survey are the following:

a. Since entering college, the proportion of students who have:

- worked full-time
- participated in study abroad
- participated in an internship program
- enrolled in an honors program or graduate level courses
- worked on a professor's research project
- assisted in teaching a course

b. Compared with when you entered college as a freshman, how would you describe your progress in:

- analytical/problem-solving skills
- ability to think critically
- writing skills
- public speaking skills
- ability to work cooperatively

c. How many undergraduate courses have you taken that emphasize:

- writing
- understanding numerical data
- scientific inquiry
- historical analysis
- foreign language skills

d. How well do the following describe the college:

- easy to see the faculty outside of office hours
- students have contact with one another outside of class - faculty are rewarded for being good teachers

e. Time Diary--number of hours per week spent in:

- class

- studying
- working

f. Frequency of involvement in:

- independent research projects
- group projects
- tutored another student
- took an essay exam
- gave a class presentation

College Student Experiences Questionnaire (CSEQ)

While not currently administered to a valid national sample, this questionnaire is particularly tailored toward the kinds of items that would be useful in building indirect indicators of "good practice." It is widely administered to college students at all types of institutions. Among the items of interest are the following:

a. Time Diary--numbers of hours per week spent in:

- school-related activity
- work-related activity

b. Levels of involvement ("quality of effort") in:

- library
- campus cultural facilities
- athletics
- other campus experiences

c. Types of courses taken in:

- writing
- math/sciences
- communications
- foreign languages

d. Frequency of key learning activities:

- thought about practical applications of material
- participated in a classroom discussion
- worked on a paper where I had to combine ideas and material from many sources
- tried to explain the material to another student
- did additional reading on assigned topics
- tried to find flaws in arguments that were made in assigned readings or classroom discussions
- student/faculty contact (several items)

e. Experiences related to writing:

- spent at least five hours writing a paper

- wrote a paper in which I had to inform, persuade, or entertain my instructor
- talked to an instructor who had made comments on a paper I had written

f. Experiences related to science/math:

- tried to express a relationship in mathematical terms
- solved an everyday problem (not a course assignment) using what I learned in science or math
- explained a scientific procedure to a friend or classmate

g. Estimated learning gains in:

- writing clearly and effectively
- knowing how to evaluate scientific claims
- putting ideas together to see relationships, similarities and differences between ideas
- developing the ability to work on my own
- developing the ability to inform, persuade or entertain others with my speaking
- understanding how numbers can be misused
- understanding how hypotheses and theories are formed, tested and validated

UCLA/HERI Faculty Survey

This survey is suitable for administration to a national sample and has been used in several national studies. Items are particularly promising in developing "good practice" indicators. Among the items of interest that the survey contains are the following:

a. Faculty activities in the past two years:

- team-taught a course
- participated in teaching-oriented workshops

b. Time Diary -- numbers of hours per week spent in:

- scheduled teaching
- preparing for teaching
- advising and counselling students
- research
- committee work
- [others]

c. Reactions to perceived campus priorities in:

- providing a quality undergraduate experience
- [other activities]

d. Reactions to perceived campus climate/environment:

- interest of faculty in the academic problems of undergraduates

- opportunities for faculty/student contact
- faculty reward structures for teaching and advising
- opportunities for student contact with one another

e. Personal objectives in teaching:

- develop ability to think clearly
- [others]

f. Evaluation methods used to assess student performance:

- multiple choice exams
- short-answer
- term papers
- final and mid-term essays
- student evaluation of each others' work
- student oral presentations
- [various types of grading]

g. Teaching methods commonly used:

- extensive lecturing
- cooperative learning
- multiple drafts of written work
- experiential learning
- extensive use of graduate teaching assistants
- group projects
- independent projects
- student-developed assignments

Review of Papers for NCES Workshop on Goal Five:
Assessing Thinking and Communication
in College Graduates

Robert Calfee
Stanford University
November 21, 1991

INTRODUCTION AND OVERVIEW

This memo reports on three papers prepared for the November workshop: Ewell and Jones Actions matter; Lenth Context and policy requisites, and Venezky Literacy. The memo begins with background on my approach to the review, followed by a summary and critique, and ends with a section on other issues and recommendations that occurred to me during the review process.

The three papers take different approaches and contain different substance. Given the criteria promulgated by NCES, I have focused on those elements with most direct relevance to the specifics of an assessment program. In my recommendations, I urge the workshop to give greater emphasis on writing as a primary indicator, to weigh the use of portfolio approaches as an assessment tool, and to rely on informed teacher judgment for evaluation and reporting of outcomes.

BACKGROUND

Two segments of the September 16 NCES project memo provide the background for my review. In the covering note, the goals of review are listed as (a) establishing the reliability and validity [sic] of the position papers; (b) identifying additional issues; and (c) framing the workshop agenda. The attachment on "Evaluation Criteria" includes one general point -- conceptual soundness -- as of primary importance. Five detailed criteria are also listed: (a) identifiable outcomes; (b) validity; (c) value added, (d) methods for accurate and informative assessment; and (e) practicality.

Taking these criteria as a whole, it seemed to me most important to speak to the pragmatics of post-secondary assessment as related to Goal Five: Ability of college graduates to think critically, communicate effectively, and solve problems in the workplace and in the practice of citizenship. A second theme, less clear in my reading, had to do with the "validity" of the position papers in framing the issues. I am not sure that the conditions of the task are adequate to support this rather daunting challenge. An adequate answer to Goal Five might require followup of graduates in the workplace and in citizenship activities for several years after

graduation -- I doubt that the political drive behind Goal Five is sufficient to support genuinely "valid" proposals of this sort. At the other extreme, it seems unlikely that the workshop will focus on development of multiple-choice tests of "basic literacy skills" to be mandated upon all college graduates.

Somewhere in the mid-range of these possibilities are techniques that can generate useful information not easily subject to manipulation and misinterpretation. My perspective in this review has been to explore such possibilities in the three papers, and to add a few thoughts from my own experience with assessment -- which ranges from kindergarten through the evaluation of teacher candidates, from research on standardized tests to exploration of informal assessment methods in classroom settings.

SUMMARY AND CRITIQUE OF PAPERS

Ewell and Jones, Indirect measures, begin with the caution that direct, performance-based indicators of college students' achievements may require an extensive and time-consuming development effort. In the short run, they argue for indirect or proxy indicators of "good practices" -- graduation requirements, student and faculty surveys, and transcript studies. The paper does not attempt any portrayal of direct studies. For the proposed indirect indicators, matters such as trustworthiness, evaluation methods, and standards, do not readily apply, in the authors' opinion. They propose instead an enhancement of existing techniques of institutional review as an interim approach.

An ancillary argument in the paper is the importance of linking student outcomes to educational inputs -- a radical idea! It's easier to attribute failure to students when no one knows what was taught and how it was taught -- NAEP has made a few efforts to include curriculum and instructional practices in its indicators, but indices of student background are more typical.

The conceptual framework for the paper in Figure 1 lays out a matrix of direct and indirect/proxy measures for student outcomes and instructional processes. The figure makes clear that current plans focus on "direct" measures of student outcomes, without little attention to other existing sources of information about either outcomes or inputs.

My main problem with the paper comes from an examination of Figure 1 (Figure 2 is hard to follow). The "stuff" that goes in the boxes is weak. Part of the problem lies in defining what is meant by the curricular substance. Where in the undergraduate course of study are students likely to gain experience in critical thinking and effective communication?

Should we conduct a semantic search of course descriptions or syllabi? Should students be required to take courses with "thinking" and "communication" in the titles -- content-free courses? I would hope not, although this approach appears the trend in some community colleges. Should we prepare surveys that ask students whether they are thinking and communicating regularly? For young people accustomed to lectures and term papers, the questions may have little meaning. I am not familiar with all of the examples that Ewell-Jones use to illustrate innovative practice, and so there may be more here than meets my eyes.

The issue comes to a head for me in the transition between pages 8 and 9 -- just when I thought the authors were going to lay out a model of instructional practices that promote thought and communication, they switch suddenly to a discussion of statistics and statistical analysis. They imply at the bottom of page 9 that existing information (the foundation for their proposal) needs "critical review and careful field testing." Why not address the assessment issues directly, rather than spending another year or two in "study." As I will argue at the end of the paper, it appears that some colleges and universities have been doing just this for a decade or more.

The authors return to meatier issues on page 11, where they list several categories of indirect indicators: proficiencies, projects, instructional contexts (e.g., class size!), and student behaviors (self reports). The impression is that most colleges do not collect such information on a regular basis, and so this "primary" information would require new efforts.

By page 15, the text reports on the types of data that are regularly or irregularly collected by post-secondary institutions. The <u>best</u> of these examples has considerable potential, in my judgment. But having spent four years as a committee chair at Stanford trying to refine and regularize a survey of this sort (and having failed), I am somewhat skeptical about the prospects for this endeavor. Why would a faculty invest significant time and energy in such activities? The concept is quite compelling, but why should students and teachers take the task seriously?

On page 18, the authors consider the strengths and limits of their proposal. I think this section might be developed into an argument for the use of their indirect indicators as the "real stuff:"

> <u>Indirect indicators are straightforward and cheap</u>: Sounds like a good beginning, assuming that the claim is sound. And I think it probably is. The measures, though sketchily presented, sound more viable than the notion that students will take seriously a standardized test with no consequences

other than to "inform Goal Five."

Indirect indicators are easy to collect: Seems likely. As to the caveat that they might be difficult to interpret. So what's new? If they are cheap and easy, then it's feasible to obtain multiple measures, which is a good tradeoff.

Indirect indicators lend guidance for improving practice: Again, a reasonable point. As to the caveat that the public may not believe them, the cost-benefit relation here is straightforward. The public "believes" that elementary and secondary students are doing better or worse based on minute fluctuations in standardized test scores. Is it good policy to commit ourselves to chasing chimera -- good politics, perhaps, but poor policy, in my opinion. The authors are more sanguine than I that "information" will provide the motivation to improve practice (p. 20), but certainly trustworthy feedback is one ingredient for improved practice.

Indirect indicators provide locally relevant information: A real strength of the proposal. To be sure, they don't necessarily lead to simple evidence about national goals. My bias in this regard is probably already evident, so I will limit myself to the observation that studies of public schooling are beginning to show the power of local initiative and the limits of externally mandated efforts.

Indirect indicators may be acceptable at first, but then opposed by local institutions: I'm not sure I agree. Seems to me that a mandated requirement to produce data will meet with resistance from the beginning. But if the movement engaged a significant number of institutions and organizations to pursue self-study and review, then a different reaction would result. At Stanford, for instance, the proposal for commitment to improved undergraduate education has caused a few "yowps," but the overall reaction has been positive. It helped that the proposal included support for the additional effort -- I suspect that Goal Five, like the others, is to be "cost free."

"What should happen next?" The paper ends with plain language. The primary recommendation is to find out what's already happening. I cannot claim expertise in the field of post-secondary indicators, but my sense is that this step should be relatively straightforward. Various organizations must have a pretty good idea of what is and what isn't available. The critical question to me is whether it makes sense to use information from the sources proposed in this chapter as indicators of college competence in problem-solving and communication.

The authors spend little time explicating how they interpret Goal Five. I like their proposal. Its weaknesses are that (a) the "goal" is not clear, and (b) the proposal may appear to be a delaying action. I am reminded of an experience when I was in basic training. Our platoon went to the firing range, and at the end of the first round my target was riddled with bulls eyes -- the soldier next to me was remarkably accurate, but had aimed at the wrong target!

A concluding thought -- the NCES guidelines ask for comment on the contribution of an assessment system to the "value added" by a college education. Seems to me that this index is critically important, and should be a significant part of self-monitoring by a responsible university. The issue is not addressed in this paper, nor in any of the others.

OTHER ISSUES

The Shopping List

Let me go back to the workshop's "shopping list," to review what seems to have been covered and neglected from the set of desiderata:

Identifiable outcomes: Ewell-Jones and Venezky both address this issue, although from quite different perspectives. My inclination is to support the Ewell-Jones proposal, but with

the addition of the kind of information covered by Venezky's analysis of literacy in the proxy measures. On the other hand, the papers leave largely untouched the operational definitions of literate thinking, problem-solving, and communication in generalizable contexts.

Defensible validity: I don't find this issue covered adequately in any of the papers. The missing ingredient was noted above, the creation of a conceptual framework for supporting the identification of significant outcomes.

Measurement of value added: Is college worth it? The baccalaureate has value in its own right, but what about the increase in competence in the literate use of language for thinking and problem solving? I happen to find this question to be of extraordinary interest and challenge. The thinking in all three papers, so far as I can tell, examines entry criteria (how to ensure that a student meets certain standards before admission to post-secondary education) and exit performance (how to measure competency on graduation). Left untouched is the connection between these two assessments, and the translation of differences into "growth" that can be related to curriculum and instruction. Most of my work is on the early years of reading acquisition, where the connections are relatively straightforward and the changes quite dramatic. I looked in vain for a similar portrayal across the four years of college.

Definable methods: Ewell-Jones and Venezky give sketchy outlines of assessment methods, quite different in character, reflecting differences in the outcomes that they proposed. Both proposals are a long way from providing an assessment design, however.

Practicality: An interesting puzzle. Practical for whom? At what cost? If the focus is on policy audiences and administrators, then practical is cheap, easily implemented and summarized, and minimally intrusive. Multiple-choice tests fill this bill quite well. If the lens switches to instructional staff, practicality means activities that support instruction, that lend themselves to assignment of grades and feedback to students, and that do not increase staff burdens. For students, practical means support for progress toward course completion and successful job finding, with prompt feedback, and without undue additional cost in time and effort. None of the papers dwells on this complex array of issues, although Lenth suggests that programs not meeting these vague criteria will sooner or later disappear.

A Troubling Trio

As I read through the various materials and papers, reflecting on my role as a college instructor, searching memory's cobwebs from early days as an undergraduate teacher, three specters kept emerging. First, what is the purpose of this endeavor, and who is the audience for the results. At one level, the answer is obvious. Goal Five must be addressed to meet the needs of policy makers to ensure the public about the quality of college education. Elementary teachers may do what they are told, but college professors (and students) are more inclined to demand justification; they ignore intrusions that they do not understand or consider ill-advised. Audience and purpose need clarification for Goal Five to get off the ground.

The second concern is process and evidence. What data are to be collected, and how are the data to be framed to make a case? It is difficult to standardize the college experience, and probably inappropriate. Here Venezky's concern about the relation of assessment to the student's degree area comes into bold relief. It seems silly to assume that "thinking and communication" are the same for the computer scientist, the English major, and the communications graduate.

The third complex centers around the issue of criteria and standards. Assessments almost always have a bottom line; a value is assigned to an object, or one object is declared of more value than another. What shall be the dimensions of evaluation? How will comparisons be drawn? And at what level (states, institutions, instructors, students?). Universities and colleges already wrestle with these issues in the certification process. Goal Five suggests that the present efforts are somehow inadequate. Less clear is the formulation of the new path.

The Importance of Writing and Interacting

The most informative lens on a person's thinking is the composing task; the most instructive data on communication skill comes from interactions in a group setting. These claims can be argued: "Why not rely on correlated proxys;" "Might be true, but it's impractical". Nonetheless, I am not convinced that the papers give significant attention to assessment of these capacities. Venezky mentions writing (not the same as composing, to be sure), and Ewell-Jones make reference to writing assignments. I think the issues are critical.

Specifically, let me suggest that valid assessment of Goal Five will depend on performance-based indicators of "real things" in contexts appropriate for post-secondary graduates. How well do college seniors write when assigned substantive projects?

How well can they "read to write;" that is, how adequately can they digest a document (one or more) and transform the contents to address a specific problem? How well do they handle extemporaneous presentations in group settings (composing without writing)? How well can an individual student work with a team of colleagues in approaching a common task?

"Writing to a prompt" in standardized settings is well traveled ground; that's not what I have in mind. Maybe I've not tuned to the right channels (nor have I perused all of the documents that appeared recently in my mail), but it seems to me that serious treatment of the questions posed above has received relatively little attention by educational assessment researchers in recent years. Reading-to write research has a small niche. "Thinking on your feet" has no literature with which I am familiar. Studies of teamwork probably reached their heyday during the Second World War (current work may be classified). But the older literature on these matters suggests that they are amenable to investigation.

I am not proposing "years of study" on these topics. Rather, the type of "think aloud" represented in Venezky's article could be extended to provide richly instructive possibilities. The Ewell-Jones "proxy" approach could cover the same range. To be sure, university professors might be startled at the request to assess students competence in group work, but they might consult colleagues who teach local kindergartens.

The Newest Fad: Portfolios

While immersed in this review, I had occasion to read the volume by Belanoff and Dickson on Portfolios: Process and product. I was surprised to discover that the work focused on post-secondary education. Portfolios are the latest fad in pre-secondary schooling, and I expected chapters extolling the virtues of collecting samples of student writing. In fact, the chapters describe several "success stories" in which writing portfolios have served to enhance curriculum, instruction, and assessment in post-secondary education. Unlike the current discussions in elementary schooling, the authors deal substantively and seriously with questions of validity, reliability, criteria for evaluation, standards of performance, and the task of reporting aggregate outcomes. The institutions range widely over small community colleges and large research universities. Many of the reports extend over a decade or more.

What comes across in this material is the centrality of teacher engagement in the design and operation of portfolio methods. In most of the case studies, the driving force was the commitment of instructors to improving the educational opportunities. To be sure, top-down pressures were important

in some instances, and the process has served as a lever for faculty development and institutional enhancement. Nonetheless, in no instance did it appear that a "standardized portfolio" was the primary vehicle for change. I urge the workshop to examine the conceptual and empirical possibilities laid out in the Belanoff-Dickson volume -- and the citations, which overlap virtually not at all with the three papers that I have reviewed.

Lessons

Policy makers who are well educated (but who may not necessarily be expert in the nature of schooling) must be frustrated by the task of mandating educational improvement. These people have traversed the path from kindergarten through post-graduate study. What's the big deal? Efficiency, standards, accountability, standardization -- establish clear goals, monitor outcomes, reward success and punish failure.

Goal Five is laudable -- the nation needs assurance that possession of a college diploma ensures competence in the literate use of language for thinking and communicating. At an earlier time, policy makers (such as there were) depended for reassurance on participants in the instructional program: professors, teaching assistants, counselors, and students.

National institutions have taken a beating in recent decades, and distrust is infective. My hunch is that the most valid touchstone for ensuring the achievement of Goal Five is the same as in earlier times. The workshop may search for ways to clarify the meaning of Goal Five, may enliven the discussion about the significance of this goal, and may even promote the development of more adequate documentation and reporting of relevant data. Efforts to centralize the fundamental process are unlikely to have much success, in my opinion.

It must be frustrating to have to trust millions of people to "do the right thing."

POST SCRIPT

The conference is concluded -- these thoughts are appended after my experiences hearing the broad ranging discussion by the participants and the reactions of the DoEd/NCES staff.

"The train is on the track" -- this remark, which emerged in a couple of versions throughout the session, seems to fit a lot current events. "We've got to do what we've got to do" is another variation from recent happenings. On the one hand, it is certainly apparent that the nation is in the doldrums, and

that change is in the air. On the other hand, a listing of the top-ranked post-secondary institutions world-wide would probably include a disproportionate number from this country. The intention of "Goal 5/Objective 5" is important, but it is also important to be clear about the aim of the effort and the effect of various actions. "Do no harm" is good advice for everyone, not just physicians. A train is one thing; a steamroller is another.

"We need your help" -- another staff comment that occurred more than once during the session. Let me reinforce this point. Public schools are (unfortunately) timid institutions. Resilient in their core, they are nonetheless accommodating on the surface. "Tell us what you want us to do" is a not uncommon reaction from students, teachers, and administrators in the K-12 field. Collaboration (working together) seldom springs from this cowed response, but neither does it lead to a battle.

Universities and colleges may be a different matter. To be sure, we are "wimpish" on the surface. We weigh and ponder, we look at all sides of the question, we dilly and dally. On the other hand, we have a fairly strong conviction that we are doing some things right -- including the task of teaching young men and women to think and to communicate. Nor are we in the "basic skills" business -- our aim is to teach our graduates to think and to communicate about **something** in a manner appropriate to the topic.

We have a history, and we are inherently conservative. Some groups may be upset by the "liberal tendencies" of college inhabitants (professors and students), but the nature of the college experience has not changed greatly over centuries. Whatever we do in the future, innovation and outcomes are likely to be measured in decades rather than years.

My sense is that the conservativism comes from a concern that, while it is easy to destroy a university, it is very difficult to create one. George Stewart of Berkeley wrote a novel some years back, <u>Earth abides</u>, that captures this theme.

And so we need to work together -- "we" at the university and anyone else who is interested in the endeavor. Imposed change is inimicable to the spirit of the free university that has flourished in this country for more than two centuries.

"The test" -- on Tuesday afternoon, as I shuttled out to Dulles from the Metro, I thought to myself, "What's the worst thing that might come from this particular initiative?" I settled on one image -- the extension of NAEP methods to college graduates. While this proposal emerged briefly in our group, and while its appeal to policy makers is apparent, I urge those in power to reject any such proposal. First, it won't really

work. Students will reject the imposition. Professors will battle the idea. Policy makers won't learn anything from the exercise. And it will use up scarce dollars for no good end.

Second, an excellent alternative exists. That is the work on portfolio assessment that is up and running on a number of college campuses throughout the country. The aim here is not to assess graduates -- they have specialized. Rather, the endeavor is to ensure that college students have attained fluency in "critical literacy" -- the capacity to use language as a tool for thinking and communicating -- before they enter upper division. The experiences reported in Belanoff and Dickson show what can be achieved for improving assessment and instruction through the portfolio approach. To be sure, the approach is not standardized, it is not subject to effective external control, and it poses challenges in aggregation. But is seems to work.

And so -- bottom line. The initiative addresses an important issue for federal policy, but it also moves the federal government into new territory. Avoid a test. Listen to the client. Try to foster critical thinking and effective communication....

U.S. Department of Education - NCES
STUDY DESIGN WORKSHOP
on
Higher Order Thinking and Communication Skills of College Graduates
November 17-19, 1991

Position Paper
"Actions Matter: The Case for Indirect Measures in Assessing Higher Education's Progress on the National Education Goals"

Author
Peter T. Evell
Dennis P. Jones

Reviewer
Elinor M. Greenberg

Overview

Upon initial reading of this paper, this reviewer was tempted to reject the case made for using "indirect indicators" as the basis, at least in the short run, for our national assessment system relative to higher order thinking and communication skills.

There were four basic reasons for this initial response. 1) The business community and the public are not in the mood to accept as credible institution-based reports of progress, good practices or innovation, based on their prior negative experiences with such institutional self-assessments. 2) Most institutions have not shown themselves to have been inclined to collect and report substantive and useful outcomes data that could be helpful to employers, graduates, or to the public. 3) Student self-reports of educational activities and their impacts are often viewed as inaccurate, vague, somewhat self-serving and, therefore, inherently flawed. My fourth reason for initial skepticism stems from my activities with NCHEMS in the 1970's and early 1980's relative to various extensive outcomes studies, which seem to have had relatively little impact on higher education, despite their validity, cost and extensive effort.

These are harsh criticisms, I know, but better that they come now from within higher education than later from outside the academic community. Upon re-review of the paper, however, my conclusions have changed, somewhat. Why?

The paper is both thoughtfully prepared and pragmatic in approach. It makes a good case for using what is already available in order to influence policy, at least in the short run, while keeping an eye on improvement as the driver for assessment efforts, in the long run. If, as the authors propose at the outset (Abstract), a few indirect measures are used concurrently with the development of other efforts, the institutions are more likely to "buy in" earlier and may see the overall national assessment task as more friendly and doable. The immediate availability of data sources is the main advantage of his approach, while the inferential nature of those sources is the main disadvantage, relative to assessing higher order thinking and communication skills.

However, given the development of a multiple option assessment plan, (as suggested by this reviewer), which could be implemented over time (1992-2000, eight years that cover two complete 4-year baccalaureate cycles), the potential "benchmark" nature of the Evell/Jones proposal could be used to launch the longer term assessment process, as other efforts follow. In this reviewer's proposed multiple option scheme, the indirect approach described in this paper is referred to as the "Institution-Based Assessment Option."

Taking a TQM "continuous improvement" approach (which should guide the entire national assessment process), indirect measures could provide important baseline information about the current state of affairs to which improvements in practice can then be linked. Additional performance-based assessments (called

by this reviewer "Development-Based," "Industry-Based" and "State-Based") could then be built on the early assessment efforts which have already begun and are likely to proceed, in any case. This combination of multiple approaches could act as a positive reinforcement to the kinds of education/industry assessment partnerships now being called for and developed, nation-wide. Then, cooperative arrangements between the Department of Education and the Department of Labor could provide the national structural framework, using data now available to each Department. This kind of inter-agency collaboration would begin to build a partnership type of structure at the federal level that could be very useful well into the 21st century and would be a model for partnerships in the field.

My guess is that within 2-5 years, these initially discreet efforts would converge, giving us a more coherent national view of performance while, at the same time, building a national consensus about the competencies needed for the new workforce in a global marketplace and how to achieve them. "Good practices" would need to be more widely publicized in order to be understood, accepted and, hopefully, utilized. At many levels, we already know what is wrong, but we've been unwilling to do the inconvenient work it will take to fix things.

This indirect, institution-based approach, then, is a conservative and pragmatic one, not designed to do the whole job of national assessment and improvement, but intended to start the process without undue delay and controversy.

Revolutions don't occur overnight. The Evell/Jones indirect approach is a first step, "necessary, perhaps, but not sufficient." The case is made well in this paper. Let's begin. We have no time to waste.

Useful Measures

Mentioned in this paper are a number of currently available useful measures that, if reported effectively and publicly in a national assessment and improvement context, would give us a starting point: GRE results, the Adult Literary Component of NEAP, the NEH analysis and the Chickering-Gamson work on "good practices," to name just a few. (There are also data in the NCES Household Survey on Adult Education, BPS, and Baccalaureate and Beyond that may be useful.)

Analysis of a current national sample of <u>transcripts</u> (NLS '72 Model), would show course-taking patterns, especially if linked to gender, ethnicity and age.

Analysis of <u>SAT writing sample data</u> could net benchmarks for college entry level writing skills of young adults which may also be useful to employers.

Reviews of <u>syllabi</u> may not be very useful, alone; but a new collection of student <u>self-reports</u> as to the "real" curriculum they experience could be compared with the curriculum as envisioned and actually taught by the faculty (K.P. Cross on the "three curricula"). This kind of comparison based on interviews in a selected sample of institutions of all types, could demonstrate and help us to learn more about current "good practices," especially if such a comparison were also linked to age, gender and ethnicity.

Data on <u>institutional requirements</u> could show what institutions might do to improve the probability of reaching the national goals and help to produce the higher order thinking and communications competencies identified as critical to improved workforce productivity and citizenship. (Note: There is a body of thought and research that points to process, not curricular content, as the key to higher order and communication skills. To be discussed further at the workshop.)

However, some new evaluations of actual <u>student/graduate achievement and performance</u>, linked to the identifiable "<u>good practices</u>" that assumedly produced them, would still have to be made, if certain practices were to be more widely accepted as causal than they are now, i.e. active learning, faculty-student contact, integrative capstone experiences, small group study, etc. The most

convincing data on the value of these practices are likely to come from students, graduates and employers, not from institutional personnel. Age, gender, ethnicity and employment status are especially relevant, here, and should be reflected in the data analyses and reports.

This paper cites a number of already available data sources which shed light on institutional practices, i.e. NCRPTAL, CSU/Vandement, NCES, Zemsky, Radcliffe, Gamson and Poulsen, UCLA, Pascarella and Terenzini, etc. These studies may already have been analyzed by the authors (Evell/Jones). If the Study Design Workshop identifies certain criteria for institutional "good practices" that can be used immediately, then reporting on these analyses could begin now.

In addition, the findings of a new eight month study, now underway, of 16 institutions and 16 (?) companies, commissioned by a leading educational association with a leading consulting firm, is due to be available early in 1992. The focus of this study is to identify the competencies seen as critical for entry-level members of the workforce relative to the global marketplace and to analyze institutional practices which contribute to developing those competencies. Findings from this study should also be factored into this "indirect" component of the national assessment strategy.

The authors are correct in stating that "this diversity of data sources is itself important in developing a reliable set of indicators." (p. 17). This is a similar argument as that used by this reviewer for designing a multiple option assessment system.

Many of the strengths and weaknesses of an "indirect" approach are spelled out in this paper. More information about Governor Ashcroft's proposed "Lifelong Learning Teams" would be helpful for our discussion in order to link our efforts to those of the NGA (1991).

The discussion of "what should happen next?" cannot be readily evaluated until after some agreement is reached as to the entire effort, not just the "indirect" measures portion. Should this "indirect" approach be the only approach used, the suggested steps seem reasonable, although not yet thoroughly flushed out as a detailed strategy or work plan. This reviewer prefers to determine next steps in a larger context. As the authors state, the "indirect" approach "seems to be a path worth exploring further" (p. 25).

Comments by Reviewer

At the risk of being repetitive, let me state here that the Evell/Jones "indirect" approach holds some real promise, if it is seen as only one early component of a decade-long effort. One real danger is that its conventional, second party, institution-based perspective will not drive either public clarity or institutional change. After all, the studies mentioned do, in fact, now exist and some data have been available for a long time. Still, major changes in institutional practices have not taken place. A number of elements are still missing in order to catalyze change.

This reviewer believes that a "perspective transformation" is required, now, in order for major change to take place. This is not easy to produce. Urgency, not patience, is the order of the day. The credibility of higher education to help increase workforce productivity in the global marketplace, as well as the quality of our citizen-based democratic process, is, indeed, at stake. Therefore, I urge us to adopt a more dramatic, multiple option, collaborative far-reaching strategy of which "indirect" measures, or an institution-based approach, is but one option among many, and one that can begin immediately.

We might think about this entire national assessment effort as a decade-long movie. This "indirect" portion is a series of snapshots that only tells us part of the story, but it holds too still to capture the true dynamic of lifelong

learning. At the same time we are looking at the frame that shows us what is happening now, we must run the reel forward, if we are to catch sight of a better future.

Conclusions

This paper makes a cogent argument for "action now" using "indirect" measures. However, this approach is not enough, alone, to satisfy the critics of higher education and may be seen as "more of the same" institutional politicking by many. While we might begin here, we cannot stop here. We must be willing to be more direct in our assessment strategy and more self-critical in our attitude. And, we must design the national assessment system in collaboration with other stakeholders (Department of Labor, the business sector, NGA, etc.) so that we are all "reading from the same page" as to desired outcomes, while maintaining the peculiar diversity of American higher education. This is also an opportune time to advance new partnerships, to make use of concepts that have recently gained broad currency, i.e. TQM, and to use the broad interest in assessment as a lever to produce significant institutional change.

Having reviewed only three papers in preparation for our Study Design Workshop, I am increasingly convinced that developing a "Collaborative Multi-Option National Assessment and Partnership System" is a valuable idea. In order to do this, the Department of Education, and especially NCES, would be required to play a consultative and coordinating role in the development of a large "Network Organization" of educational providers. This would require some shift in role and strong leadership in communicating that shift, along with its value and timeliness, to institutions, other federal departments, Congress, the President, the Governors, the states, and the public.

This is our challenge, as I see it. This is the potential promise that our Study Design Workshop holds. I look forward to our discussions and appreciate the opportunity to participate.

Review of Actions Matter: The Case for Indirect Measures in Assessing Higher Education's Progress on National Education Goals

Mary L. Tenopyr

In the initial pages of this paper, the authors suggest that the expected evaluation criteria cannot be applied to the approach of indirect measurement they propose. Consequently, it will be difficult to examine the proposals in as structured a manner as might be desired. However, an effort will be made to frame the evaluative comments presented here in such a manner that at least the major criteria are satisfied.

This paper presents approaches involving surveys, self description questionnaires, transcript studies, existing assessment results in such a way as to obtain indirect assessments of higher order thinking and communication skills.

The authors recommend multiple measures the results of which will be combined after research is completed. They recommend this approach, because they contend that it can be done rapidly and the data are relatively easy to obtain.

Relative to the primary evaluation criteria, I have the following observations:

A) The writing is not so concise as might be desired. Since the approach suggested is in itself exploratory, it would be difficult to present it in a concise manner and still persuade the reader that it was worthwhile.

B) The reasoning is as straight-forward as it can be, considering the nature of the proposal, but there appear to be a number of flaws:

1) There was no mention of the effect of relationship between initial competence of entering students and the kinds of interventions that would be effective. Nor is there mention of the interaction of students' skills with the quality of the institution.

2) The inaccuracy of self ratings of ability is well known, yet somehow the authors imply that students may be able to do a task requiring even finer discrimination. That is, estimating gains in skills.

3) The argument that obtaining the indirect measures could be done rapidly is problematic. It is this reader's opinion that arriving at an acceptable set of indirect measures may take as long as developing tests.

4) I believe that the authors underestimate the difficulty of establishing causal linkages and actual gains, when no satisfactory direct measurement that may serve as a criterion is available.

5) Research on life history data has indicated that some of the logically most valid experiences are not those empirically most valid.

C) Considering the rather exploratory nature of the proposal, the arguments are well framed, if not entirely persuasive.

D) The authors appear to have done a literature review that is commensurate with the short paper, but unfortunately the nature of the proposal is such that the first steps will be to seek the supporting information.

In conclusion, I believe that the authors have an approach worth considering, but it cannot be thoroughly evaluated until considerably more work is done.